

Data-driven geophysical forecasting: Simple, low-cost, and accurate baselines with kernel methods

Boumediene Hamzi¹, Romit Maulik^{*2}, and Houman Owhadi³

¹Department of Mathematics, Imperial College London, London, SW7 2AZ, UK.

²Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont, IL 60439, USA.

³Department of Computing and Mathematical Sciences, Caltech, Pasadena, CA 91125, USA.

Abstract

Modeling geophysical systems as dynamical systems and regressing their vector field from data is a simple way to learn emulators for such systems. We show that when the kernel of these emulators is also learned from data (using kernel flows, a variant of cross-validation), then the resulting data-driven models are not only faster than equation-based models but are easier to train than neural networks such as the long short-term memory neural network. In addition, they are also more accurate and predictive than the latter. When trained on observational data for the global sea-surface temperature, considerable gains are observed by the proposed technique in comparison to classical partial differential equation-based models in terms of forecast computational cost and accuracy. When trained on publicly available re-analysis data for temperatures in the North-American continent, we see significant improvements over climatology and persistence based forecast techniques.

1 Introduction

The numerical simulation of geophysical processes for forecasting is limited by high computational costs, parameterization requirements across various domains, and the spatial/temporal resolution of fast processes. Data-driven methods are becoming increasingly popular for surrogate modeling (emulation) or data analyses of all or a portion of the processes involved in numerical weather prediction [18, 11, 26, 29]. These include deep learning, and neural network methods [34, 33, 28, 22], which have shown exceptional performance at the cost of a loss of interpretability. We combine interpretable kernel methods with kernel selection methods (Kernel Flows [19], as presented in [10] for learning dynamical systems) for the forecasting of geophysical data sets and show competitive results at exceptionally low computational costs for sub-seasonal temperature forecasting when the underlying kernels are also learned from data. In particular, we demonstrate that these methods provide interpretable and computationally efficient alternatives¹ to deep learning architectures, which are infrastructure hungry and require off-nominal assessments for model interpretability. The framework proposed here is amenable to physics emulation, a-posteriori error estimation, input and model form uncertainty-quantification while being computational inexpensive from a learning perspective. The goal of this study is *not* to claim superiority over other geophysical emulators but to advocate using simple data-driven methods as effective baselines before migrating to compute hungry deep learning methods.

2 Related work

Temperature forecasting is a crucial capability for several applications relevant to agriculture, energy, industry, tourism, and the environment. Improved accuracy in short and long-term forecasting of air and sea-surface temperature has

^{*}Corresponding author: rmaulik@anl.gov

¹Trained neural networks themselves can also be viewed as kernel machines [6, 9] or warping kernels regressors [20].

significant implications for cost-effective energy policy, infrastructure development, and downstream economic consequences [5]. The current state of the art in temperature forecasting is obtained with partial differential equation (PDE) based methods [27, 16]. Since these methods generally require solving large systems of equations with high-performance computing resources, they are limited by access and considerations of energy-efficiency.

This area has recently become a popular application of machine learning (ML) methods due to the promise of comparable (if not greater) forecast accuracy at a fraction of the computational cost. This also enables uncertainty quantification through ensemble forecasts, which are impossible for PDE-based methods due to their excessive computational complexity. For these reasons, there has been a great degree of interest in building ML ‘emulators’ or ‘surrogate models’ from various geophysical data sets. There have been several studies on the use of machine learning for accelerating geophysical forecasts in recent times. Several rely on using machine learning methods to devise parameterizations for processes that contribute a significant cost to the numerical simulation of the weather and climate [4, 8, 1, 23, 17]. Other studies have looked at complete system emulators (i.e., forecasting from data alone) with a view to forecasting without any use of and consequent limitations of equation based methods [14, 28, 35, 24, 3, 25]. Other studies have looked at utilizing historical information for forecasting specific processes using data from the process alone [30, 15, 31]. Opportunities and perspectives for the use of data-driven methods for the geosciences may be found in [12, 7].

In this article, we introduce an entirely data-driven method for forecasting the maximum air temperature over the North-American continent. We achieve this by obtaining a low-dimensional affine subspace approximation of the temperature field on which a reduced system is evolved. Both dimensionality reduction and system evolution are performed using data-driven techniques alone, with the former employing a proper-orthogonal decomposition and the latter combining kernel methods with a cross-validation technique known as Kernel Flows. We remark that in contrast to the growing popularity of deep learning methods for forecasting, we propose using classical dimensionality reduction and time-series forecasting with suitable inductive biases. Our competitive results motivate the creation of simple data-driven baselines that compare favorably to PDE-based methods without specialized neural architectures.

3 Data set(s)

3.1 Weekly averaged sea surface temperature

For our geophysical emulation, we use the open-source National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation sea surface temperature (SST) V2 data set.² Seasonal fluctuations in this data set cause strong periodic structure, although complex ocean dynamics still lead to rich phenomena. Temperature snapshot data is available on a weekly basis on a one-degree grid. This data set has previously been used in data-driven forecasting and analysis tasks (for instance, see [13, 2]), particularly from the point of view of identifying seasonal and long-term trends for ocean temperatures by latitude and longitude. Each “snapshot” of data comprises a temperature field in an array of size 360×180 (i.e., the longitudes and latitudes of a one-degree resolution grid), which corresponds to the average sea-surface temperature magnitude for that week. Prior to its utilization for forecasting, a mask is used to remove missing locations in the array that corresponds to the land area. The nonzero data points then are flattened to obtain an \mathbb{R}^Z -dimensional vector as our final snapshot for a week.

This data is available from October 22, 1981, to June 30, 2018 (i.e., 1,914 snapshots). We utilize the period of October 22, 1981, to December 31, 1989, for training and validation (427 snapshots). The remaining data set (i.e., 1990 to 2018) is used for testing (1,487 snapshots). This breakdown of the data set into training and testing is commonly used [2] and the 8-year training period captures seasonal as well as subdecadal trends in the data set. The training data is utilized to obtain data points given by a window of inputs and a window of outputs corresponding to the desired task of forecasting the future, given observations of the past sea-surface temperatures. These data points are then split into training and validation sets. We note that this forecast is performed non-autoregressively—that is, the data-driven method *is not* utilized for predictions beyond the desired window size. Since this data set is produced by combining local and satellite temperature observations, it represents an attractive forecasting task for data-driven methods.

²Available at <https://www.esrl.noaa.gov/psd/>

3.2 North-American daily midnight surface temperature

The National Centers for Environmental Prediction's (NCEP) North-American Mesoscale Forecast System (NAM) [27] is one of the main mesoscale models used for guiding public and private sector meteorologists. NAM runs four times daily at three different spatial scales: (1) Full North-American 12-km resolution; (2) 4-km Continental US (CONUS) nest, 6-km Alaska nest, and 3-km Hawaii and Puerto Rico nests. These domains are one-way nested inside the 12-km domain; (3) High-resolution nested domain which has a different location each cycle based upon the NCEP Service Centers and National Weather Service Offices. For this work, analysis data from NAM was used, using the 12-km resolution grid for the surface temperature. The NAM data is collected in a time period between the 28th of October 2008 to the 20th of September 2018 on a daily cadence. In particular, we measure the temperature at midnight on each day in this temporal domain. This corresponds to 3569 snapshots of sea and land surface temperature data. In contrast to the previous problem, our goal is to forecast one week in advance with the temperature resolved daily. The first 2555 snapshots are reserved for the purpose of training the proposed forecasting technique, while the rest are used for testing. In a manner similar to the previous test case, a time-delay of 7 days is used to specify the inputs to obtain the 7-day forecast of temperature in the future.

4 Methods

4.1 Dimensionality reduction: Proper orthogonal decomposition

Proper orthogonal decomposition (POD) provides a systematic method to project dynamics of a high dimensional system onto a lower-dimensional subspace. We suppose that a single snapshot of the full system is a vector in \mathbb{R}^N , where N could be the number of grid points at which the field is resolved. Observing the system across a number of time points gives us the snapshots $\theta_1, \dots, \theta_T$, with mean subtracted by convention. The aim of POD is to find a small set of orthonormal basis vectors v_1, \dots, v_M , with $M \ll N$, which approximates the spatial snapshots,

$$\theta_t \approx \sum_{j=1}^M a_j(t) v_j, \quad t = 1, \dots, T, \quad (1)$$

and so allows us to approximate the evolution of the full N dimensional system by considering only the evolution of the M coefficients $a_j(t)$. POD chooses the basis, v_j , to minimize the residual with respect to the L_2 norm,

$$R = \sum_{t=1}^T \left\| \theta_t - \sum_{j=1}^M a_j(t) v_j \right\|^2. \quad (2)$$

Defining the snapshot matrix, $S = [\theta_1 | \dots | \theta_T]$, the optimal basis is given by the M eigenvectors of SS^T , with largest eigenvalues, after which, the coefficients are found by orthogonal projection, $a_j(t) = \langle \theta_t, v_j \rangle$ [32].

For both of our data sets, we take only the training data snapshots, say D_1, \dots, D_T , from which we calculate the mean $\bar{D} = (1/T) \sum_t D_t$, hence defining the mean subtracted snapshots $\theta_t = D_t - \bar{D}$. We then create the snapshot matrix, S , and find numerically the M eigenvectors of SS^T with largest eigenvalues. From this, we train models, \mathcal{N} , to forecast the coefficients

$$(a(t+1), a(t+2), \dots, a(t+K)) \approx (\hat{a}(t+1), \hat{a}(t+2), \dots, \hat{a}(t+K)) = \mathcal{N}(a(t), a(t-1), \dots, a(t-K)). \quad (3)$$

making predictions of future coefficients given previous ones. Here, K is the duration of the embedding delay chosen as a hyperparameter.

To test the predictions on unseen data, E_1, \dots, E_k , we take the mean \bar{D} , and vectors v_j calculated from the training data to get that

$$a_j(t) = \langle E_t - \bar{D}, v_j \rangle, \quad j = 1, \dots, M, \quad (4)$$

which will be used by the model \mathcal{N} to make a prediction for future coefficients. The prediction for the coefficients \hat{a} , can be converted into predictions in the physical space by taking $\bar{D} + \sum_j \hat{a}_j v_j$. This procedure only makes use of testing data to pass into the model, not to train the model in any way. Crucially, to make a forecast of $E_{t+1}, E_{t+2}, \dots, E_{t+K}$, only previous measurements E_t, E_{t-1}, \dots are needed.

4.2 Cross-validation for temporal forecasting

The method employed for temporal forecasting was introduced in [10] for learning dynamical systems from data. It is a kernel ridge regression with a kernel learned from data from a variant of cross-validation known as Kernel Flows [19].

To describe this method let x_1, \dots, x_k, \dots be a time series in \mathbb{R}^d (d is, in our application, the number of modes used for prediction). Our goal is to forecast x_{n+1} given the observation of x_1, \dots, x_n . We work under the assumption that this time series can be approximated by a solution of a dynamical system of the form

$$z_{k+1} = f^\dagger(z_k, \dots, z_{k-\tau^\dagger+1}), \quad (5)$$

where $\tau^\dagger \in \mathbb{N}^*$ and f^\dagger may be unknown. Here, in our application, each z_k is time-window of POD coefficients, i.e. of the form $z_k = (x_{k-\tau'}, \dots, x_k)$ where τ' is the length of the forecast window ($\tau' = 8, 7$ in our respective data sets). Given $\tau \in \mathbb{N}^*$ ($\tau' = \tau$ in our experiments), the approximation of the dynamical system can then be recast as that of interpolating f^\dagger from pointwise measurements

$$f^\dagger(X_k) = Y_k \text{ for } k = 1, \dots, N, \quad (6)$$

with $X_k := (x_{k+\tau-1}, \dots, x_k)$, $Y_k := x_{k+\tau}$ and $N = n - \tau$. For the experiments in this study, forecasts of length τ are concatenated during testing to obtain predictions in the entire testing timescales.

Given a reproducing kernel Hilbert space of candidates \mathcal{H} for f^\dagger , and using the relative error in the RKHS norm $\|\cdot\|_{\mathcal{H}}$ as a loss, the regression of the data (X_k, Y_k) with the kernel K associated with \mathcal{H} provides a minimax optimal approximation [21] of f^\dagger in \mathcal{H} . This interpolant (in the absence of measurement noise) is

$$f(x) = K(x, X)(K(X, X))^{-1}Y, \quad (7)$$

where $X = (X_1, \dots, X_N)$, $Y = (Y_1, \dots, Y_N)$, $k(X, X)$ for the $N \times N$ matrix with entries $k(X_i, X_i)$, and $k(x, X)$ is the N vector with entries $k(x, X_i)$. This interpolation has also a natural interpretation in the setting of Gaussian process (GP) regression: (i.) (7) is the conditional mean of the centered GP $\xi \sim \mathcal{N}(0, K)$ with covariance function K conditioned on $\xi(X_k) = Y_k$, and (ii.) the interpolation error between f^\dagger and f is bounded by the conditional standard deviation of the GP ξ , i.e.

$$|f^\dagger(x) - f(x)| \leq \sigma(x) \|f^\dagger\|_{\mathcal{H}}, \quad (8)$$

with

$$\sigma^2(x) = K(x, x) - K(x, X)(K(X, X))^{-1}K(x, X)^T. \quad (9)$$

Evidently, the accuracy of the proposed approach depends on the kernel K , and, as in [10], we also learn that kernel from the data (X_k, Y_k) with Kernel Flows (KF) [19].

Given a family of kernels $K_\theta(x, x')$ parameterized by θ , the KF algorithm can then be described as follows [19, 36]:

- i. Select random subvectors X^b and Y^b of X and Y (through uniform sampling without replacement in the index set $\{1, \dots, N\}$)
- ii. Select random subvectors X^c and Y^c of X^b and Y^b (by selecting, at random, uniformly and without replacement, half of the indices defining X^b)
- iii. Let³

$$\rho(\theta, X^b, Y^b, X^c, Y^c) := 1 - \frac{Y^{c,T} K_\theta(X^c, X^c)^{-1} Y^c}{Y^{f,T} K_\theta(X^b, X^b)^{-1} Y^b}, \quad (10)$$

be the squared relative error (in the RKHS norm $\|\cdot\|_{K_\theta}$ defined by K_θ) between the interpolants u^b and u^c obtained from the two nested subsets of the dataset and the kernel K_θ

- iv. Evolve θ in the gradient descent direction of ρ , i.e. $\theta \leftarrow \theta - \delta \nabla_\theta \rho$

- v. Repeat.

In the following, we denote the use of RKHS to forecast POD coefficients as the POD-RKHS emulation framework. We will be using a kernel that is a combination of the triangular kernel, locally periodic kernel, polynomial, gaussian, Laplace and other generic kernels.

³ $\rho := \|u^b - u^c\|_{K_\theta}^2 / \|u^b\|_{K_\theta}^2$, with $u^b(x) = K_\theta(x, X^b) K_\theta(X^b, X^b)^{-1} Y^b$ and $u^c(x) = K_\theta(x, X^c) K_\theta(X^c, X^c)^{-1} Y^c$, and ρ admits the representation (10) enabling its computation

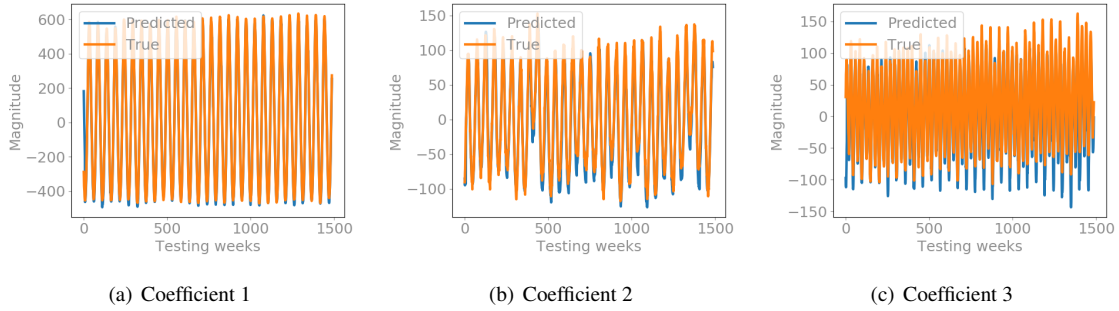


Figure 1: POD Coefficient predictions for the NOAA-SST data set using POD-RKHS. We utilize the time period of October 22, 1981, to December 31, 1989, for training and validation (427 snapshots). The remaining data set (i.e., 1990 to 2018) is used for testing (1,487 snapshots) and is shown here.

5 Experiments

In the following section, we outline results from the use of the POD-RKHS framework for forecasting on the aforementioned data sets. We also provide metrics that evaluate the accuracy of the forecasts and compare them with baseline techniques. Our first set of results are shown for the NOAA-SST data set. The ability to forecast weekly averaged coefficients eight weeks at a time is shown in Figure 1, (for the testing period), where it can be observed that the lower order structures are predicted with high fidelity. The framework shows deviations in the finer scale content (mode 3) as one approaches the end of the testing period indicating potential extrapolation. Similar behavior was observed in [15, 31] as well, where a deep LSTM was used to forecast in this reduced space (henceforth POD-LSTM). Time series assessments for various point probes in the Eastern Pacific are shown for a testing sub-window (where all data is available) (to add time period here) in Figure 2. The plot indicates a competitive testing performance when compared to state-of-the-art equation-based methods such as the Community Earth System Model (CESM) and U.S. Navy Hybrid Coordinate Ocean Model (HYCOM). The former is a climate simulator, and the latter is an operational forecast system giving short term predictions. Both these techniques require vast computational resources, whereas our forecast is performed on a single node machine without any accelerator.

Some qualitative comparisons of the predictions are shown in Figure 3 where an acceptable agreement between different methods and the remote-sensing data set is observed (see Table 1). A closer examination of the mean squared error during the entire testing period, shown in Figure 4 where POD-RKHS is seen to give sufficiently accurate predictions in the entire domain, including the vital Eastern Pacific region. With the testing period being around 18 years, the El-Niño Southern Oscillation (ENSO) oscillation, occurring at a frequency of 2-7 years [29], is captured more accurately in comparison to CESM. In terms of times to solution, data-driven models provided forecasts for the given time period (1981–2018) almost instantaneously. This can be contrasted with CESM (for the forecast period of 1920–2100), which required 17 million core-hours on Yellowstone, the National Center for Atmospheric Research (NCAR) high-performance computing resource, for each of the 30 members of the large ensemble. While compute costs in such high detail are not available for HYCOM, this short-term ocean prediction system runs daily at the Navy Department of Defense Supercomputing Resource Center, with daily data typically accessible within 48 hours of the initial run time⁴. Benchmarking for the 1/25 degree HYCOM forecasts (twice finer than the reference data used here) indicates the requirement of 800 core-hours per day of forecast on a Cray XC40 system.⁵ For a more appropriate comparison, we also take into account the computational cost for the search of an optimal LSTM architecture and its training. While the POD-RKHS method requires approximately 40 seconds to train on a single node machine without acceleration, the LSTM being compared with was discovered using 3 hours of wall time on 128 compute nodes of the Theta supercomputer with the Intel Knights Landing architecture. POD-RKHS was trained only once on one Intel CoreI7 x86_64 CPU without any acceleration. The best LSTM was retrained on our data set for a longer duration of approximately 20 seconds on the same single node machine. Subsequent assessments on a different data set (in the next set of experiments) indicate that the RKHS continues to be competitive in training costs.

Our second set of assessments are performed for the NOAA-NCEP NAM data set for sea and land surface air

⁴<https://www.hycom.org/dataserver>

⁵https://www.hycom.org/attachments/066_talk_COAPS.17a.pdf

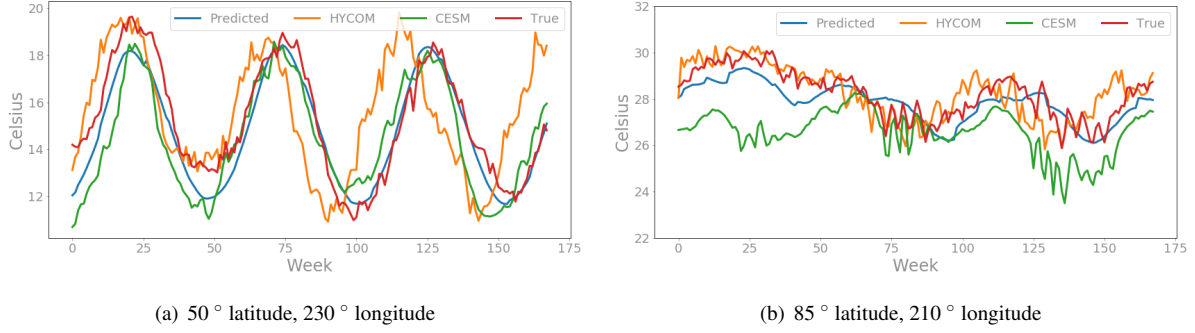


Figure 2: Probe time-series comparisons for NOAA-SST from CESM, HYCOM and POD-RKHS within the testing regime at two different locations. The data are plotted for weeks between April 5, 2015, and June 17, 2018.

Table 1: RMSE breakdown (in Celsius) for different forecast techniques compared against the POD-LSTM forecasts between April 5, 2015, and June 24, 2018, in the Eastern Pacific region (between -10 to +10 degrees latitude and 200 to 250 degrees longitude). POD-RKHS matches the accuracy of the process-based models for this particular metric and assessment and that of an optimized LSTM [15] using neural architecture search.

	RMSE ($^{\circ}$ Celsius)							
	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
POD-LSTM	0.62	0.63	0.64	0.66	0.63	0.66	0.69	0.65
CESM	1.88	1.87	1.83	1.85	1.86	1.87	1.86	1.83
HYCOM	0.99	0.99	1.03	1.04	1.02	1.05	1.03	1.05
POD-RKHS	0.76	0.67	0.66	0.69	0.69	0.72	0.77	0.76

temperatures over the North-American continent. This data set is obtained by a re-analysis of numerical simulation data obtained through data assimilation of observations. We follow the same strategy of time-delayed inputs to outputs for forecasting the midnight temperature. Figure 5 shows the ability of the time-series emulator to extract an underlying trend from a highly noisy signal.

Note that the difficulty of extracting the extremely high-frequency features contributes to the poor capture of extreme events during emulation. The reconstruction accuracy from the forecast is compared in a series of assessments beginning with root-mean-squared error (RMSE) assessments as shown in Figure 6 for the testing time period. POD-RKHS is seen to provide competitive results in comparison to persistence and climatology at the lower latitudes. In contrast, exceptional gains are seen at the higher latitudes with far reduced RMSE values from the proposed framework. Note that errors across all frameworks were maximum in the Northern section of the continent, proving the increasing difficulty of predicting in that region using data-driven techniques. Figure 7 plots the contours for improvements in the correlation coefficient and cosine similarities when compared to climatology. While correlation coefficients indicate regions where the predictions improve over the baselines in an averaged sense, Cosine similarities indicate the ability to detect extreme fluctuations in a forecast. The contours indicate that the proposed framework improves on extreme fluctuation detection over climatology in the north region of the spatial domain as well. Similar trends are observed for comparisons with persistence with large gains in Cosine similarities in the Northern part of the spatial domain.

For the purpose of comparison, we also outline the performance of a standard LSTM framework (manually selected in comparison to [15]) for forecasting on the NAM data set. The problem formulation is identical, with the LSTM utilizing a window of inputs of 7 days length and forecasting the POD coefficient trajectories for the next 7 days. We train 55 instances of this LSTM, consisting of 3 stacked LSTMs and 50 neurons for each of the linear operations in each cell, and select the one with the best training and validation performance. The 55 LSTM architectures were trained using an Nvidia V100 GPU and required 5 hours of wall time. For each training, an early stopping criterion is utilized to terminate training if validation losses do not improve for ten successive training epochs. Figure 8 shows predictions from the testing data using the trained LSTM with clear indications of poor capture of dynamics. These are confirmed by contour plots for the RMSE, shown in Figure 6, where the POD-LSTM framework is the weakest of all the data-driven methodologies presented here. Figure 9 examines potential improvements in the correlations

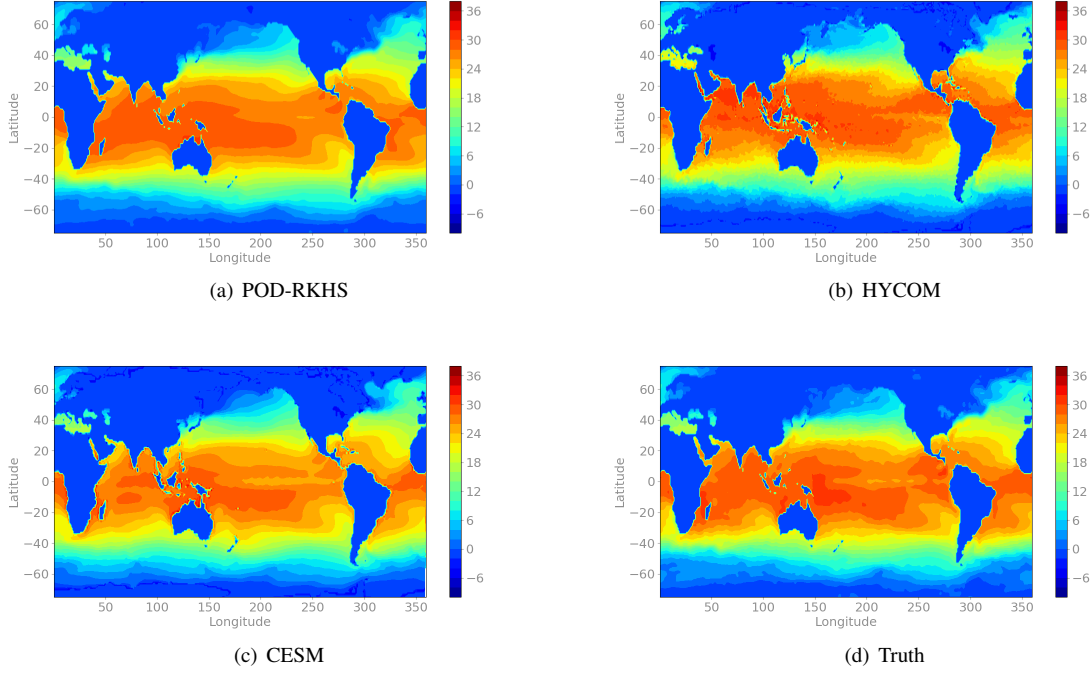


Figure 3: Contour plots for various forecasts (a-POD-RKHS, b-HYCOM, c-CESM, d-Truth), indicating how low-dimensional manifold based emulation reduces the ability to capture fine-scaled features in the flow-field. HYCOM, with the finest resolution, is seen to capture small scale information most accurately. Note, however, that the POD-based emulation framework is competitive in an averaged sense, as seen through RMSE metrics.

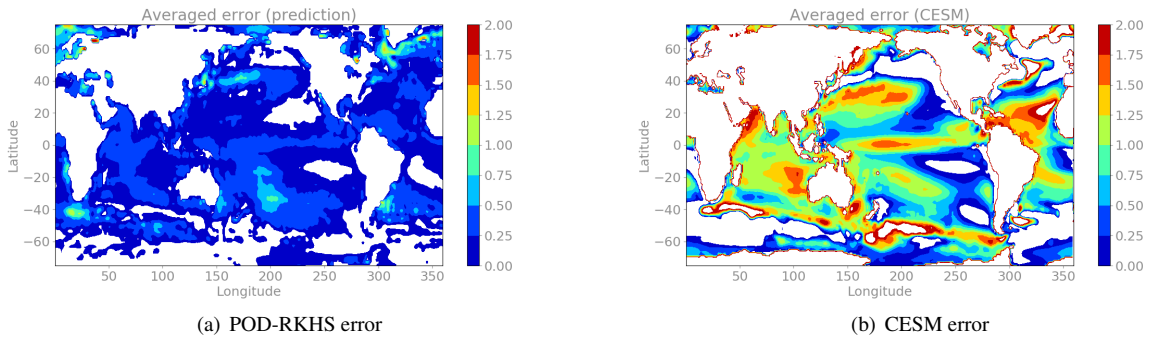


Figure 4: Contour plots for RMSE error over the entire testing period (1990-2018) for the NOAA-SST data set for POD-RKHS (a) and CESM (b). The results indicate that the proposed low-dimensional emulator captures the long term behavior of the SST fluctuations more accurately than CESM.

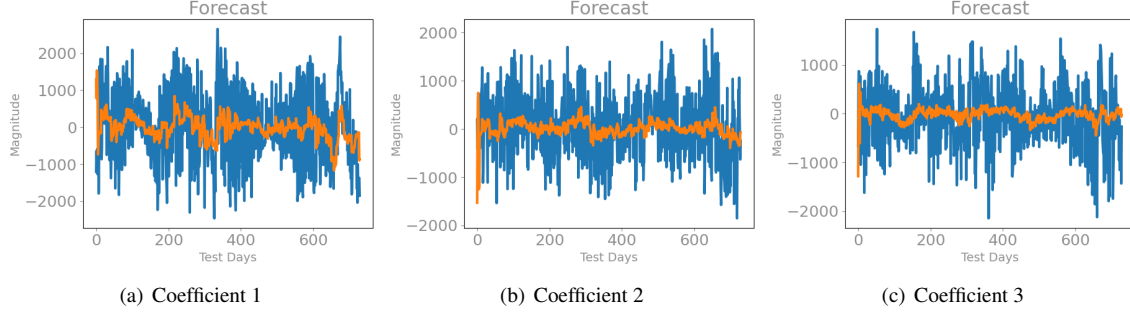


Figure 5: POD coefficient predictions for forecasting in the testing regime for the NOAA-NCEP NAM data set. The plots indicate the stochastic nature of the daily temperature at midnight and also highlight the ability of the POD-RKHS framework to extract a signal from them.

and Cosine similarities within the spatial domain of the data set for this experiment. Climatology and persistence are seen to comprehensively outperform a standard LSTM architecture. We note that this result is not meant to justify mistrust of all deep learning architectures but propose the use of classical kernel-based methods for the construction of baselines.

We now compare the training and testing times of the proposed framework with those of LSTM methods with standard architectures. The POD-RKHS method required a total of 136 seconds to train, whereas the POD-LSTM methods took an average of 330 seconds to train. Although the choice of user-defined hyperparameters in both these methods may change these values, their training times are comparable.

6 Discussion

In this article, we have employed a simple kernel ridge regression method for the forecasting of geophysical time series. We have shown that when the kernel is also learned from data (via Kernel Flows), then the proposed framework provides competitive and computationally efficient baselines for the geophysical emulation of two temperature-based data sets given by the weekly averaged sea-surface temperature (NOAA-SST) and the daily resolved midnight temperature (NCEP-NAM). In both cases, the proposed method recovers a stable temporal behavior in a low-dimensional manifold. Importantly, for the second data set, strong stochasticity is handled elegantly by the proposed framework to obtain superior metrics when compared to climatology or persistence. Comparisons with the POD-LSTM framework, commonly used for non-intrusive reduced-order modeling, are also favorable. While an LSTM obtained through expensive parallelized neural architecture search slightly outperforms the proposed framework for the relatively smooth NOAA-SST time-series data, the proposed framework is much more successful in the case of the stochastic NCEP-NAM temperature data. Training costs of the proposed method and the LSTM are also comparable, with both providing model fits in similar training duration on single node machines without any specialized deep learning accelerators.

Extensions to this work include learning stochastic emulators and the addition of more covariates to the training data to emulate larger fractions of the earth system model. The latter may be obtained by concatenating reduced representations of the flow field (in truncated POD space) to obtain a larger, more informative state vector. Another avenue for exploration is to use kernel flows for obtaining reduced-representations themselves in a unified pipeline for geophysical emulation. Both these endeavors are instrumental for pushing data-driven forecast horizons to regimes where classical methods are untrustworthy.

7 Data availability

The data that support the findings of this study are openly available in Github at https://github.com/Romit-Maulik/POD_RKHS.

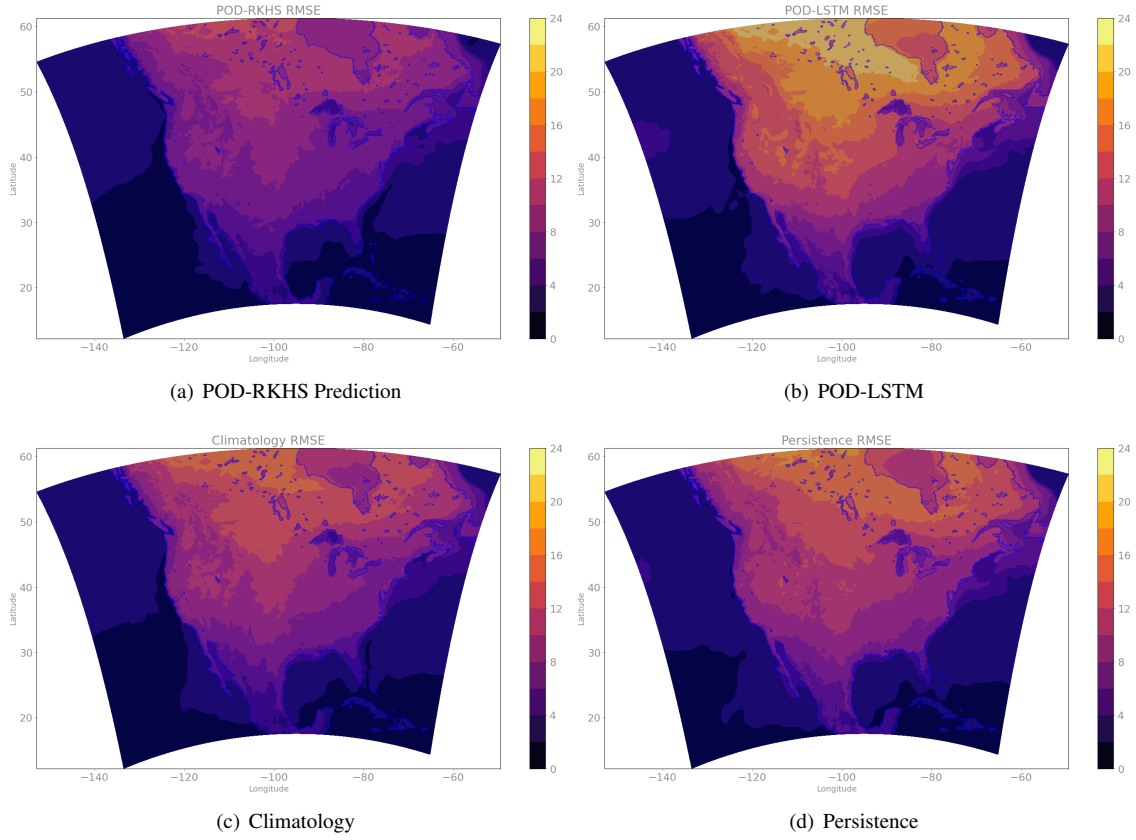


Figure 6: Testing RMSEs for predictions on the NAM data as compared to climatology and persistence baselines for the POD-RKHS and POD-LSTM frameworks. This result indicates the use of one potential LSTM with manual tuning - further improvements may be obtained through neural architecture and hyperparameter search as seen in [15].

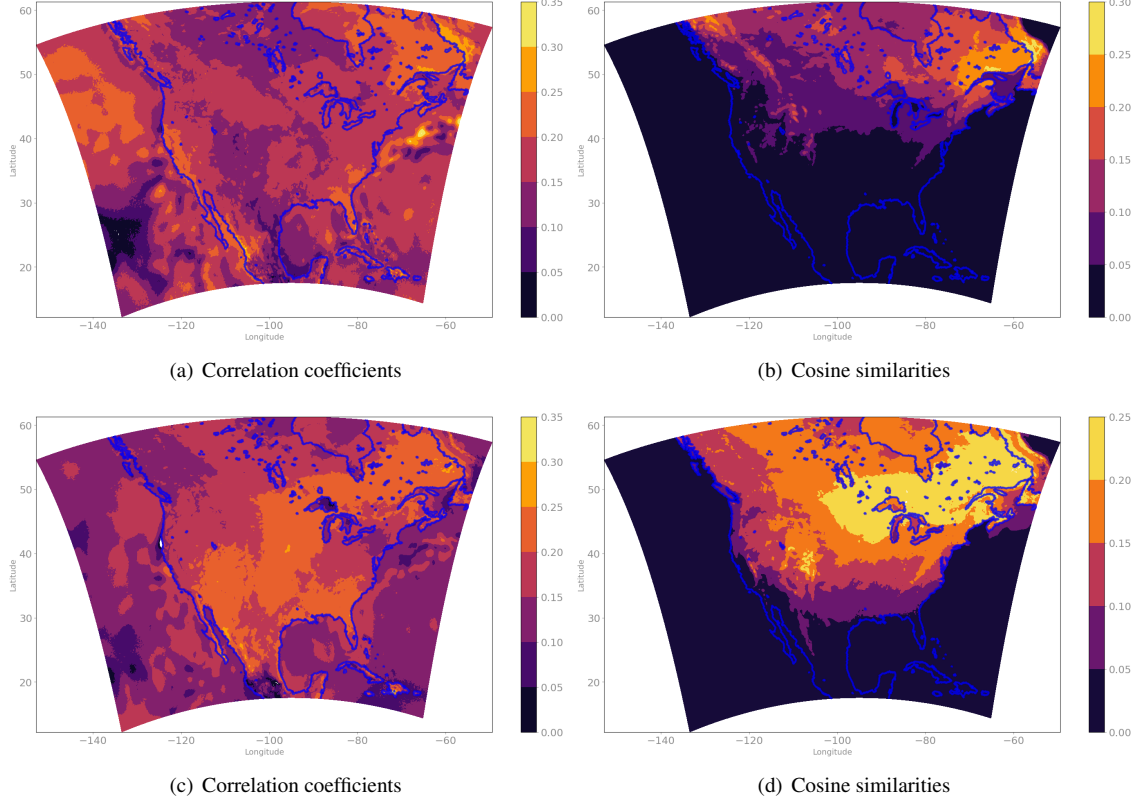


Figure 7: Metric **improvements** on climatology (top) and persistence (bottom) using the POD-RKHS time-series emulator. The north domain of the data set shows improved performance by the proposed framework, particularly for extreme fluctuation detection (indicated by Cosine similarity improvement). Overall correlation with the truth is seen to be improved throughout the domain.

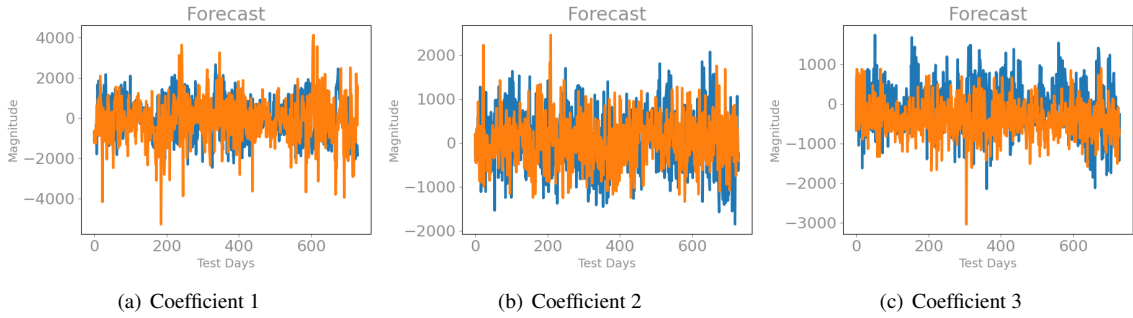


Figure 8: Predictions for the NOAA-NCEP NAM POD Coefficients using a POD-LSTM framework. This prediction was obtained with 3 stacked LSTM cells, each with 50 neurons for all the internal operations. An LSTM that provided the best training and validation performance among 55 restarts was chosen. The total time to train these LSTMs was 5 hours on an Nvidia V100 GPU. It is observed that the LSTM architecture is sensitive to the stochastic nature of the data.

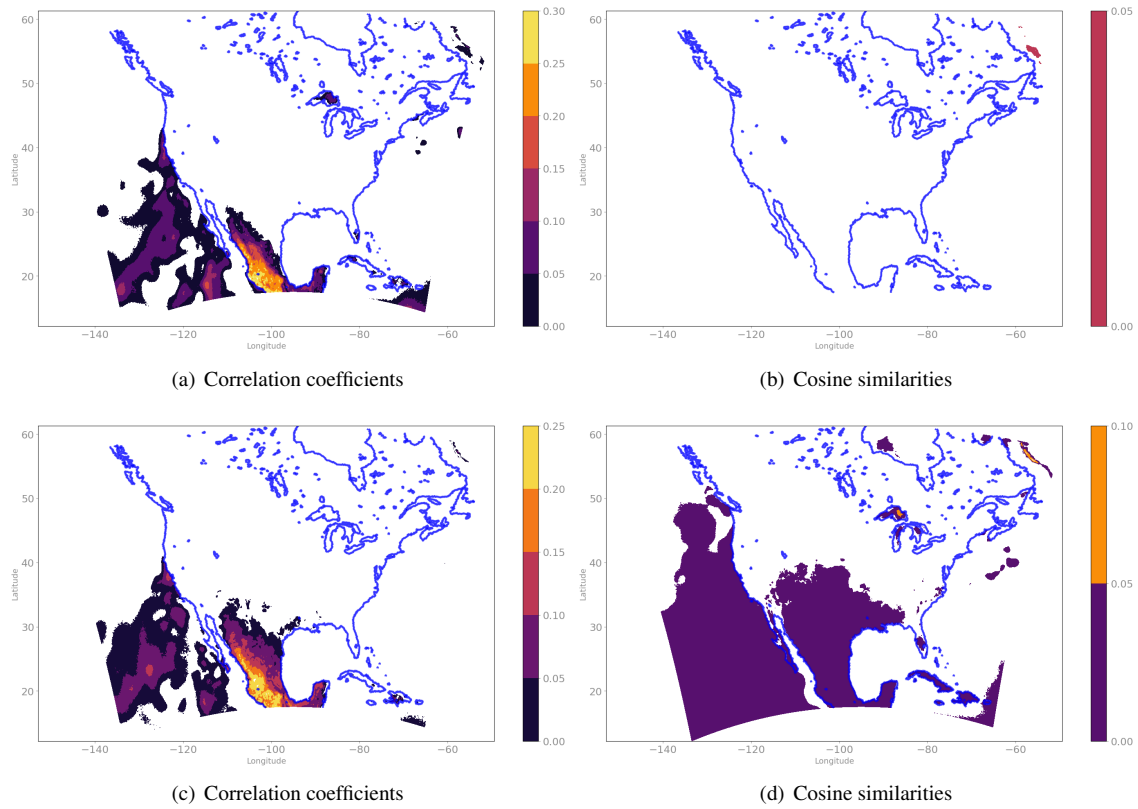


Figure 9: Metric **improvements** on climatology (top) and persistence (bottom) using the LSTM time-series emulator. Vast stretches (i.e., the white regions) of the domain show poorer performance for POD-LSTM against climatology and persistence baselines.

8 Acknowledgment

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research, under Contract DE-AC02-06CH11357. This research was funded in part and used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. R. M. acknowledges support from the Margaret Butler Fellowship at the Argonne Leadership Computing Facility. B. H. thanks the European Commission for funding through the Marie Curie fellowship STALDYS-792919 (Statistical Learning for Dynamical Systems). H. O. gratefully acknowledges support by the Air Force Office of Scientific Research under award number FA9550-18-1-0271 (Games for Computation and Learning) and MURI (FA9550-20-1-0358). This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. DOE or the United States Government.

References

- [1] Noah D Brenowitz and Christopher S Bretherton. Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, 45(12):6289–6298, 2018.
- [2] Jared L Callahan, Kazuki Maeda, and Steven L Brunton. Robust flow reconstruction from limited measurements via sparse representation. *Physical Review Fluids*, 4(10):103907, 2019.
- [3] Ashesh Chattopadhyay, Ebrahim Nabizadeh, and Pedram Hassanzadeh. Analog forecasting of extreme-causing weather patterns using deep learning. *Journal of Advances in Modeling Earth Systems*, 12(2):e2019MS001958, 2020.
- [4] Ashesh Chattopadhyay, Adam Subel, and Pedram Hassanzadeh. Data-Driven Super-Parameterization Using Deep Learning: Experimentation With Multiscale Lorenz 96 Systems and Transfer Learning. *Journal of Advances in Modeling Earth Systems*, 12(11):e2020MS002084, 2020.
- [5] Melissa Dell, Benjamin F Jones, and Benjamin A Olken. Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, 4(3):66–95, 2012.
- [6] Pedro Domingos. Every model learned by gradient descent is approximately a kernel machine. *arXiv preprint arXiv:2012.00152*, 2020.
- [7] Peter D Dueben and Peter Bauer. Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, 11(10):3999–4009, 2018.
- [8] Pierre Gentine, Mike Pritchard, Stephan Rasp, Gael Reinaudi, and Galen Yacalis. Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45(11):5742–5751, 2018.
- [9] Klaus-Robert Muller Grégoire Montavon, Mikio L. Braun. Kernel analysis of deep networks. *Journal of Machine Learning Research*, 12:2563–2581, 2011.
- [10] Boumediene Hamzi and Houman Owhadi. Learning dynamical systems from data: a simple cross-validation perspective, Part I: parametric kernel flows . *Physica D*, 2021.
- [11] Chris Huntingford, Elizabeth S Jeffers, Michael B Bonsall, Hannah M Christensen, Thomas Lees, and Hui Yang. Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters*, 14(12):124007, 2019.
- [12] Anuj Karpatne, Imme Ebert-Uphoff, Sai Ravela, Hassan Ali Babaie, and Vipin Kumar. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 31(8):1544–1554, 2018.
- [13] J Nathan Kutz, Xing Fu, and Steven L Brunton. Multiresolution dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 15(2):713–735, 2016.

- [14] Yunjie Liu, Evan Racah, Joaquin Correa, Amir Khosrowshahi, David Lavers, Kenneth Kunkel, Michael Wehner, William Collins, et al. Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *arXiv preprint arXiv:1605.01156*, 2016.
- [15] Romit Maulik, Romain Egele, Bethany Lusch, and Prasanna Balaprakash. Recurrent neural network architecture search for geophysical emulation. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '20. IEEE Press, 2020.
- [16] Franco Molteni, Roberto Buizza, Tim N Palmer, and Thomas Petroliaxis. The ECMWF ensemble prediction system: Methodology and validation. *Quarterly journal of the royal meteorological society*, 122(529):73–119, 1996.
- [17] Peter D Nooteboom, Qing Yi Feng, Cristóbal López, Emilio Hernández-García, and Henk A Dijkstra. Using network theory and machine learning to predict El Niño. *Earth System Dynamics*, 9(3):969–983, 2018.
- [18] Paul A O’Gorman and John G Dwyer. Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, 10(10):2548–2563, 2018.
- [19] H. Owahdi and G. R. Yoo. Kernel flows: From learning kernels from data into the abyss. *Journal of Computational Physics*, 389:22–47, 2019.
- [20] Houman Owahdi. Do ideas have shape? Plato’s theory of forms as the continuous limit of artificial neural networks. *arXiv preprint arXiv:2008.03920*, 2020.
- [21] Houman Owahdi and Clint Scovel. *Operator-Adapted Wavelets, Fast Solvers, and Numerical Homogenization: From a Game Theoretic Approach to Numerical Approximation and Algorithm Design*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2019.
- [22] Mr Prabhat, Jim Biard, Sangram Ganguly, Sasha Ames, Karthik Kashinath, Soo Kyung Kim, Samira Kahou, Tegan Maharaj, Christopher Beckham, Travis Allen O’Brien, et al. ClimateNet: a machine learning dataset for climate science research. In *AGU Fall Meeting Abstracts*, volume 2017, pages IN13E–01, 2017.
- [23] Stephan Rasp, Michael S Pritchard, and Pierre Gentile. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39):9684–9689, 2018.
- [24] Stephan Rasp and Nils Thuerey. Purely data-driven medium-range weather forecasting achieves comparable skill to physical models at similar resolution. *arXiv preprint arXiv:2008.08626*, 2020.
- [25] Eduardo Rocha Rodrigues, Igor Oliveira, Renato Cunha, and Marco Netto. Deepdownscale: a deep learning strategy for high-resolution weather forecast. In *2018 IEEE 14th International Conference on e-Science (e-Science)*, pages 415–422. IEEE, 2018.
- [26] David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433*, 2019.
- [27] Suranjana Saha, Shrinivas Moorthi, Hua-Lu Pan, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, Robert Kistler, John Woollen, David Behringer, et al. The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8):1015–1058, 2010.
- [28] Sebastian Scher. Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. *Geophysical Research Letters*, 45(22):12–616, 2018.
- [29] Oliver T Schmidt, Gianmarco Mengaldo, Gianpaolo Balsamo, and Nils P Wedi. Spectral empirical orthogonal function analysis of weather and climate data. *Monthly Weather Review*, 147(8):2979–2995, 2019.
- [30] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *arXiv preprint arXiv:1506.04214*, 2015.

- [31] Dominic J Skinner and Romit Maulik. Meta-modeling strategy for data-driven forecasting. *arXiv preprint arXiv:2012.00678*, 2020.
- [32] Kunihiko Taira, Maziar S. Hemati, Steven L. Brunton, Yiyang Sun, Karthik Duraisamy, Shervin Bagheri, Scott T. M. Dawson, and Chi-An Yeh. Modal analysis of fluid flows: Applications and outlook. *AIAA Journal*, 58(3):998–1022, 2020.
- [33] Benjamin A Toms, Karthik Kashinath, Da Yang, et al. Deep learning for scientific inference from geophysical data: The Madden-Julian Oscillation as a test case. *arXiv preprint arXiv:1902.04621*, 2019.
- [34] Mirko Van der Baan and Christian Jutten. Neural networks in geophysical applications. *Geophysics*, 65(4):1032–1047, 2000.
- [35] Jonathan A Weyn, Dale R Durran, and Rich Caruana. Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, 12(9):e2020MS002109, 2020.
- [36] Gene Ryan Yoo and Houman Owhadi. Deep regularization and direct training of the inner layers of neural networks with kernel flows. 2020. <https://arxiv.org/abs/2002.08335>.